Policy Research Working Paper 6364

Impact Evaluation Series No. 84

Using Provider Performance Incentives to Increase HIV Testing and Counseling Services in Rwanda

Damien de Walque
Paul J Gertler
Sergio Bautista-Arredondo
Ada Kwan
Christel Vermeersch
Jean de Dieu Bizimana
Agnès Binagwaho
Jeanine Condo

The World Bank
Development Research Group
Human Development and Public Services Team
February 2013

Abstract

Paying for performance provides financial rewards to medical care providers for improvements in performance measured by specific utilization and quality of care indicators. In 2006, Rwanda began a paying for performance scheme to improve health services delivery, including HIV/AIDS services. This study examines the scheme's impact on individual and couples HIV testing and counseling and using data from a prospective quasi-experimental design. The study finds a positive impact of paying for performance with an increase of 6.1 percentage points in the probability of individuals having ever been tested. This positive impact is stronger for married individuals: 10.2 percentage points. The results also indicate larger impacts of paying for performance on the likelihood that the respondent reports both partners have ever been tested, especially among discordant couples (14.7 percentage point increase) in which only one of the partners is HIV positive.

This paper is a product of the Human Development and Public Services Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at ddewalque@worldbank.org and gertler@haas.berkeley.edu.

Produced by the Research Support Team

Using Provider Performance Incentives to Increase HIV Testing and Counseling Services in Rwanda.

Damien de Walque*, Development Research Group, The World Bank

Paul J Gertler*, Haas School of Business, University of California, Berkeley

Sergio Bautista-Arredondo, National Institute of Public Health, Cuernavaca, Mexico

Ada Kwan, National Institute of Public Health, Cuernavaca, Mexico

Christel Vermeersch, The World Bank.

Jean de Dieu Bizimana, Camris International

Agnès Binagwaho, Ministry of Health, Kigali, Rwanda

Jeanine Condo, National University of Rwanda, School of Public Health, Kigali, Rwanda

Keywords: Performance-based financing; HIV testing and Counseling; Africa

JEL classification: I12; O15.

* These authors shared first authorship. Corresponding author: Damien de Walque, Development Research Group, The World Bank, 1818 H Street, NW, Washington DC 20433, USA. Tel: 202-473-2517, Fax: 202-614-0234, ddewalque@worldbank.org. We thank Anita Asiimwe, Paulin Basinga, Stefano Bertozzi, Gyuri Fritsche, Alex Kamurase, Kathy Kantengwa, Gayle Martin, Rigobert Mpendwazi, Cyprien Munyanshongore, Vedaste Ndahirdwa, Isaac Ntahobakulira, Miriam Schneidman, Jennifer Sturdy, Claude Sekabaraga, Louis Rusa, Adam Wagstaff, and many others for their contributions to the project; the Rwandan Ministry of Health, the National AIDS Control Commission (CNLS) in Rwanda, Management Sciences for Health, the Belgian Technical Cooperation, Cordaid, GTZ, Healthnet, USAID, and the World Bank and their staff for their collaboration in the implementation of the P4P rollout plan and supporting the evaluation; and the National University of Rwanda School of Public Health for data collection. We thank the World Bank's Bank-Netherlands Partnership Program and Spanish Impact Evaluation Fund, the British Economic and Social Research Council and the Government of Rwanda for their financial support. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1. Introduction

HIV testing and counseling (HTC) is a gateway to improving prevention and care efforts, and has become a core strategy for decreasing HIV transmission and incidence (Glick, 2005). There have been calls to devote more resources to couple HTC since HIV transmission is highest in discordant couples, i.e. couples in which only one of the partners is infected, especially if the infected partner either does not know his or her status or has not revealed it to the uninfected partner (Padian et al., 1993). Recent evidence demonstrates that antiretroviral treatment (ART) of HIV+ individuals is very effective in preventing transmission of the HIV virus within couples (Cohen et al., 2011; Dodd et al., 2010; El-Sadr et al., 2010; Wagner et al., 2010). As a result, HTC couple testing, especially among discordant couples, has become a key component of prevention programs in generalized epidemic countries.

Despite the promise of HTC to reduce HIV transmission and the large amount of development assistance for HIV/AIDS, HTC uptake has only recently seen modest improvements (United Nations, 2011). Moreover, there are few documented successful experiences of HTC programs reaching couples (Padian et al., 1993; Painter, 2001).[1]

A promising, yet largely untested, intervention to increase testing is to pay health providers for increasing participation in HTC. This is part of the more general Pay-for-performance (P4P) movement that gives financial rewards at the facility and/or provider levels to improve performance measured by specific utilization and quality of care indicators. P4P is now being piloted or scaled up in over 20 low- and middle-income countries (Eichler and Levine, 2009; Meessen et al., 2011).

The purpose of this paper is to evaluate the impact of Rwanda's national P4P scheme on individual and couple HTC. Building on the lessons from pilot experiences in a few provinces, Rwanda initiated in 2006 a national P4P scheme at the health center level to improve health services delivery, including HIV/AIDS services. We use data from a prospective impact evaluation we nested into the national scale-up of P4P in Rwanda, producing evidence from an

[1] An important exception is from Thornton (2008) who demonstrated that cash value vouchers doubled the percentage of individuals who obtained their HIV test results, given that they had been tested.

impact evaluation at scale with more external validity than closely monitored pilot experiments. The Rwanda P4P scheme provided larger payment for couple HTC than for individual HTC, allowing us to explicitly test whether supply-side incentives are an effective intervention to increase couple HTC and in particular for discordant couples among whom the risk of HIV transmission is higher.

An important aspect of our study is the identification of the incentives' effects. P4P affects health care provision in two ways: first, through incentives for providers to expend more effort in specific activities and second through an increase in the amount of financial resources available to the health center. In order to identify the incentive effect separately from the increase in resources, the traditional input-based budgets of the comparison group were increased by the average amount of incentive payments to treatment facilities. As a result, while treatment and comparison facilities had the same financial resources available, a portion of the treatment facilities' budgets was determined based on their performance whereas the comparison facilities' resources were not. This is important because if P4P achieves its results just from increased financial resources, then the same results could be achieved from a simple increase in budget without incurring the administrative costs associated with implementing the incentive scheme.

Our results show a positive impact of P4P with an increase of 6.1 percentage points in the probability of individuals having ever been tested. However, when disaggregated by couple status we find that all of the results are driven by individuals living in a couple. There is no effect on single individuals even when we condition on being sexually active. However, there is a positive and statistically significant impact of 10.2 percentage points for individuals in couples, which amounts to a 14.5% increase over the control group testing rate. The impact of P4P on couple testing is particularly strong among discordant couples (i.e. one partner is confirmed HIV+ and the other is not), encouraging the partners of identified HIV patients to come for HTC. These results are consistent with the fact that the Rwanda P4P strongly encouraged couple and partner testing, paying US$ 0.92 per new individual tested for HIV and US$ 4.59 per couple/partner jointly tested.

These results show that incentive payments are an effective means of increasing participation in HTC. They are especially important for Sub-Saharan Africa, where nearly 80% of HIV-infected adults are unaware of their HIV status and over 90% do not know whether their partners are infected (World Health Organization, 2009). With only 12% of the global population, Sub-Saharan Africa is home to 68% of all people living with HIV.[2]

Our findings contribute to the limited but growing evidence base that paying health facilities for performance is a feasible and effective method for improving health system performance in low- and middle-income countries. Our work contributes to the general literature on P4P in medical care, as it is the first to examine the impact of P4P incentives on HIV related services.[3] More importantly, the role of incentives in P4P is key. Because the comparison facilities' regular budgets were increased by an amount equal to the P4P payment to the treatment group, we were able to isolate the P4P incentive effect from the resource effect.

Our work also contributes to the relatively small literature on the effects of paying medical care providers for performance in developing countries.[4] There are four well-identified and related evaluations in other low- and middle-income countries. Hospital-based physicians in the Philippines who received extra bonus pay based in part on knowledge of appropriate clinical procedures reported increases in clinical knowledge (Peabody et al., 2011).. In Indonesia, performance incentives to villages for improvements in health outcomes led to an increase in labor supply from health providers (Olken et al., 2012). Miller et al. (2012) found that bonus payments to schools significantly reduced anemia among students in China. Finally, using the same identification strategy as this study, Basinga et al. (2011) found in Rwanda that P4P had significant positive impact on institutional deliveries and preventive care visits by young children, and improved quality of prenatal care, but found no effect on the number of prenatal care visits or on immunization rates. A follow-on study also reported large impacts on child health outcomes and provider productivity (Gertler and Vermeersch, 2012).

[2] In 2011 an estimated 34 million people were living with HIV worldwide, the number of AIDS-related deaths was 1.7 million and there were 2.5 million new HIV infections (UNAIDS, 2012).

[3] See Witter et al. (2011) for a recent systematic review of health care performance incentives in low- and middle-income countries. Most of the literature that they cite do not have control groups and estimate the impact of P4P as jumps in time trends of the amount of services providers by treatment facilities.

[4] There is, however, a growing literature on P4P for medical care in the U.S. and other high-income countries with mixed results. See (Alshamsan et al., 2010; Scott et al., 2011; Van Herck et al., 2010).

The remainder of the paper is organized as follows. Section 2 describes the context of the health sector in Rwanda and the P4P intervention evaluated. In section 3, we present our data and we describe our identification strategy. Section 4 presents our results while section 5 concludes.

2. The health sector in Rwanda and the P4P intervention

In 2005, HIV prevalence for adults in Rwanda was estimated at 3% (Institut National de la Statistique du Rwanda (INSR) and ORC Macro, 2006). The Government of Rwanda (GoR) decided to address the HIV epidemic by not only aggressively scaling up HIV services nationwide, but also utilizing the planned national P4P model to target HTC, ART for AIDS patients, prevention of mother-to-child transmission (PMTCT), and other HIV-related prevention and care services. The GoR initiated the P4P scheme in 2006 to supplement the input-based budgets of health centers and hospitals with bonus payments based on the quantity and quality of key services (Ministère de la Santé République du Rwanda, 2006).

The scheme pays for different dimensions of services, including maternal and child health, tuberculosis, and HIV/AIDS. For HIV/AIDS, the P4P scheme pays for 10 output indicators, such as the number of clients and the number of couples tested for HIV (US$ 4.59 per couple), the number of newly diagnosed HIV-positive patients on ART (US$ 0.92 per individual), and the number of HIV-positive women on contraception (Table 1). The Ministry of Health (MoH) defined the output indicators and each corresponding unit payment based on health priorities, available budget and the previous NGO pilot experiences (Ministère de la Santé République du Rwanda, 2008). This analysis focuses on the first two indicators dealing with HTC: (i) the number of clients tested for HIV at the HTC center, and (ii) the number of couples/partners tested at the health facility.

Facilities submit monthly reports and quarterly requests for payment to the district P4P steering committee, which is responsible for verifying the data and authorizing payment. Each committee verifies reports by sending auditors to facilities on unannounced random days each quarter. The auditors verify the data reported are the same as the data recorded in facility records. In addition, during the 2006-2008 period the Ministry of Health financed one patient tracking survey to conduct face-to-face interviews with approximately 1,000 patients to verify the accuracy of the records. This survey found false reporting was below 5 percent (Health Development and Performance, 2008).

Quarterly payments go directly to facilities and are used at each facility's discretion. In the sample of 10 treatment facilities in our study, the P4P payments amounted to 14 percent of overall expenditures in 2007. On average, facilities allocated 60 to 80 percent of the P4P funds to increase personnel compensation.

It is worth noting that the Rwanda P4P scheme was implemented in the context of a larger health sector reform and during a period in which HIV/AIDS services, including delivery of antiretroviral treatment, were extensively scaled-up. As of 2005, 83 health facilities were delivering ART to 19,058 persons living with HIV/AIDS (PLWHA), and 229 facilities were providing HTC services with 449,259 individuals ever tested (Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics, 2007). By 2008, coverage of ART had increased more than threefold and more than doubled for HTC (Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics, 2008).

3. Impact evaluation design, data and baseline characteristics

3.1 Impact evaluation design

The evaluation design with sample sizes is presented in Figure 1. The evaluation design took advantage of the phased implementation of the program at the district level. Administrative districts with pre-existing NGO P4P pilot schemes were excluded from the sampling frame. The remaining districts were then grouped into eight blocks based on similar characteristics for rainfall, population density, and predominant livelihoods using data from the 2002 Census blocks included between two and four districts depending on characteristics and size. Facilities within each district within each block were then randomly assigned into the treatment group and the other into the comparison group.

Prior to implementation of the baseline survey, the administrative district boundaries were redrawn in the context of a decentralization effort (MINALOC (Ministry of Local Government), 2006). As a result, some of the experimental areas were combined with areas that already had the NGO run pilot P4P schemes. Because P4P could not be "removed" from health facilities that were already implementing the system, and because P4P was managed at the district level, the GoR required that all facilities within those new districts be in the first phase (treatment) of the rollout. This led the evaluation team to switch the assignment of treatment and comparison for eight districts from four blocks, as well as add one block to the sample. In the end, the study's nine blocks include 16 districts in total, of which 9 belonged to the treatment group and 7 to the comparison group. Facilities that had P4P prior to the study were excluded from the evaluation (Vermeersch et al., 2010). Only facilities offering HIV/AIDS services were included in the HIV/AIDS P4P model. The 24 facilities that offered HIV services in these districts were enrolled in the study, with 10 facilities in the treatment and 14 facilities in the comparison group.

Facilities in the treatment group started receiving the HIV/AIDS related P4P in January 2007, while the facilities in the comparison group continued with traditional input-based financing until July 2008. Exposure to the HIV P4P model varied between 12 to 15 months in treatment facilities.

Since a primary objective of the evaluation was to isolate the P4P incentive effect separately from the effect of an increase in financial resources, it was necessary to hold the level of resources constant across treatment and comparison facilities. To accomplish this, comparison facilities' traditional input-based budgets were increased by the average amount of P4P payments to treatment facilities on a quarterly basis during the experiment. Therefore, the differences in outcomes between the two groups at follow-up must be attributed to the difference in incentive structures and not to a difference in available financial resources.

3.2 Data

We conducted a baseline survey of the facilities from August until November 2006 and a follow-up survey from April until July 2008.[5] We also designed a household level survey that

[5] The study team submitted the research protocol to the Rwanda National Ethics Committee, which approved the research design, methodology and methods for informed consent. Because of the sensitive nature of the sample and survey, the data collection was managed by the National University of Rwanda School of Public Health with guidance from the National AIDS Control Commission (CNLS). The follow-up surveys were also reviewed by the National Institute of Public Health in Mexico Institutional Review Board (IPF Code 3627801).

was administered to a sample of 1,000 households with an HIV+ positive member, and 600 randomly sampled neighbor households in the catchment area of the facility. We identified HIV/AIDS patients either by contacting the health facility where they received care or via association of PLWHA. We selected them randomly proportional to the number of HIV/AIDS patients attending each facility. We obtained informed consent from the patients before interviewing their household. In the follow-up survey, 85 percent of the baseline households were re-interviewed. The rate of attrition from the baseline sample was not statistically different between the treatment and comparison groups (15 percent each).

The outcome measures are constructed using data from the household surveys. The outcome for individual HTC is an indicator variable for whether the individual has ever been tested for HIV. For the purpose of the HTC analysis, we exclude identified HIV/AIDS patients who by definition are HIV positive and are aware of their HIV status. The sample is further restricted to individuals aged 15 or older. At baseline, the sample is comprised of 438 individuals in the treatment group and 445 in the comparison group. Individuals present at follow-up but not at baseline were not different based on standard socio-demographic characteristics.

For the analysis of couple testing, we create an indicator using the question of whether or not the most recent sexual partners the respondents had in the 12 months prior to the survey had ever been tested for HIV. We further combine the responses about each respondent's individual testing and the testing of their sexual partners to create an indicator variable for whether both partners in the couple/sexual partnership have ever been tested. We then restrict the sample to individuals living with their sexual partners and who self-reported having had sex in the 12 months prior to the survey. For this analysis, the unit of observation is the couple and we include only one report by couple to avoid double-counting.

3.3. Summary statistics and balance at baseline

Table 2 reports the baseline means of facility characteristics in 2006. Confirming that the evaluation design achieved balance of observed characteristics at baseline between the facilities in the treatment and comparison groups, there are no significant differences between the treatment and comparison groups in terms of rural location, proportion of district hospitals in the sample of facilities, proportion of facilities that are government-assisted or public, size of catchment population, supply of staff, log 2006 expenditures, and allocation of the budget across medical personnel, medical supplies and non-medical purposes.

Table 2 further reports the mean 2008 log expenditures for treatment and comparison facilities with no statistically significant difference in the means after the introduction of P4P in the treatment facilities. This confirms that the program compensated the comparison facilities with an increase in their traditional input-based budget equal to the increase in treatment facilities' resources and validates the interpretation of any estimated impacts being caused by the introduction of P4P incentives, as opposed to an increase in financial resources.

For the analysis of HTC at the individual level, the sample consists of all adults aged 15 and above who were identified as HIV negative: 438 in the treatment group and 445 in the comparison group. Table 3 reports the baseline characteristics of all respondents grouped and by marital status[6]. There are no statistical differences in baseline means of the outcome variable "ever been tested". For the control variables used in the regression models, the samples are generally well balanced. All samples and sub-samples are well balanced in terms of sexual activity, marital status and assets values. The proportion of individuals whose partner is identified as HIV patient is also well balanced so that the proportion of discordant couples[7], i.e. a couple where one partner is identified as an HIV patient and one is not, is well balanced across treatment and control.

For the analysis of HTC at the couple level, the sample consists of all adults aged 15 and above who were identified as HIV negative, who self-reported having had sex in the 12 months preceding the survey, and living with their sexual partners: 179 in the treatment group and 180 in the comparison group. Table 4 reports the baseline characteristics of respondents. There are no statistical differences in baseline means of the 3 outcome variables: "has the respondent ever been tested", "has the sexual partner of the respondent ever been tested" and "have the respondent and his/her sexual partner ever been tested". The only difference is that, overall,

[6] For marital status, we defined as married or living in couple both those legally married and those cohabiting together even without formal marriage.

[7] Strictly speaking, a sero-discordant couple is a couple where one of the partners is HIV positive and the other is HIV negative. Since we did not test the study participants for HIV, the only participants whose HIV status is known to us are those identified as HIV/AIDS patients who are HIV positive. Hence our definition of discordant couple is different: we define as discordant couple a couple where one partner is identified as an HIV patient and one is not.

respondents in the control group are about 3 years older than those in treatment group. All other variables including education and asset value are well balanced.

3.4. Estimation

Given the reassignment of districts between the treatment and comparison groups before the start of the study, and the limited number of districts that could be assigned to the treatment and comparison groups, we view our study as quasi-experimental. While the sample is balanced at baseline on outcomes and characteristics, it is possible that the reassignment of districts was correlated with something unobservable to us and related to health outcomes. However, redrawing of administrative units took place within the context of a decentralization agenda that was led by the Ministry of Local Government, and we find no evidence that it was driven by or related to health outcomes (MINALOC (Ministry of Local Government), 2006).[8] Therefore, we think it is likely that any relevant unobservable factors were likely to be invariant over the time period of the intervention.

Therefore, we will use the difference-in-differences method that controls for unobserved time invariant characteristics.[9] This method compares the change in outcomes in the treatment group to the change in outcomes in the comparison group. By comparing changes, we control for observed and unobserved time invariant characteristics as well as for time-varying factors that are common to the treatment and comparison groups. As we discussed above, the final assignment to the treatment and comparison groups is orthogonal to pre-intervention observable variables, leading us to believe that there is likely no correlation between this assignment and unobservables that would drive program effects.

---

[8] According to MINALOC (Ministry of Local Government) (2006), the objective of the decentralization was to enhance institutional development and capacity building for responsive local governance, to develop efficient, transparent and accountable fiscal and financial management systems at local government and grassroots levels.

[9] An alternative, sometimes used in the literature, is the intent to treat estimator that compares the originally assigned treatments to controls. In this case, however, we would have misassigned 40% of the observations and would be grossly underpowered. Also, all of the examples we could find use the ITT in cases where the study entered the field intending to implement the original design and where behavioral choices by the study participants compromised the study design. In our case, the design was changed before we entered the field and was not compromised by the study participants. Hence, while our difference in difference estimator requires stronger assumptions, we believe that it is appropriate in terms of identification and is valid based on the balance tests and knowledge of the institutional context that drove the change in design. In our view, the difference in difference choice maximizes potential power without sacrificing internal validity.

We treat the 2006 and 2008 household surveys as repeated cross-sections and estimate the following regression specification of the difference-in-difference model for individual outcomes:

$$Y_{ijt} = \alpha_j + \gamma_{2008} + \beta \cdot P4P_j \cdot I_{2008} + \sum_k \lambda_k X_{kijt} + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} is the HTC outcome of individual or couple i living in facility j's catchment area in year t; $P4P_j$ is a dummy variable that takes value 1 if facility j belongs to Phase I (i.e. started receiving P4P in 2007) and 0 otherwise; α_j is a facility fixed effect; γ_{2008} is a fixed effect for 2008; I_{2008} is a dummy variable that takes value 1 if the year of observation is 2008 and 0 otherwise; the X_{kijt} are individual characteristics; and ε_{ijt} is a zero mean error term. We compute robust standard errors clustered at the district by year level to correct for correlation of the error terms across facilities within districts.

4. Results

Table 5 reports the estimated P4P program impacts on HTC outcomes using the individual as the unit of analysis. We present analyses for the entire sample and then conduct sub-group analyses by marital status. In all of the estimated models, we control for age, education, and household assets and for gender and marital status when relevant. Assets are measured as the value of land, durables in the house, farm animals, farm equipment, and microenterprise equipment.

In column (1) of table 5, we find a positive and statistically significant impact at the 10% level of 6.1 percentage points in the probability to have ever been tested with respect to the comparison group. This represents a 10.6 percent increase over the control group. When we restrict the sample to individuals living in a couple (column 2), we find a positive impact of 10.2 percentage points that is statistically significant at the 5% level, representing a 14.5 percent increase from baseline. However, there is no impact on individuals not currently in a couple regardless of whether they are sexually active or not (columns 3 and 4). Those larger impacts of

P4P on HTC among married individuals are consistent with the fact that the P4P scheme strongly encouraged couple and partner testing since it paid US$ 0.92 per new individual tested for HIV and US$ 4.59 per couple/partner jointly tested.

In Table 6 focusing on the analysis where the couple is the unit of observation, there are significant positive impacts of P4P on the likelihood that the respondent reports that both partners have ever been tested: an increase of 8.6 percentage points, significant at the 10% level, corresponding to a 12 percent increase from baseline (column 1). That increase is especially strong among couples in which one of the partners has been identified as living with HIV/AIDS (discordant couple): the results in column 2 indicate an increase of 14.7 percentage points, significant at the 5% level and representing an 18.14 percent increase from baseline. The increase is lower and not statistically significant for couples, which are not discordant (column 3).

This analysis with the couples as the units of observation confirms that the larger P4P incentives for joint testing especially encourage both partners in the couple to be tested. The impact of P4P on couple testing is particularly strong among discordant couples where one of the partners has been identified as living with HIV/AIDS, encouraging the partners of identified HIV patients to come for HTC.

5. Conclusions

Our study examines the impact of the national P4P scheme in Rwanda on individual HTC and couple HTC, using data from a prospective experimental design. The results indicate a positive impact of P4P with an increase of 6.1 percentage points in the probability of individuals having ever been tested. This positive impact is concentrated among individuals in couples. The results also indicate larger impacts of P4P on the likelihood that the respondent reports both partners have ever been tested, especially among discordant couples in which only one of the partners is HIV positive.

Our results show significant increase of HTC coverage in the context of a massive scaling-up of HIV services. P4P was implemented in the context of a larger health sector reform and during a period in which HIV/AIDS services, including delivery of ART, were extensively scaled-up. We are not able to identify how this context of increase of HIV service coverage interacted with the P4P program. Arguably a P4P intervention could have even greater impacts in a more static context of HIV service delivery.

Strong encouragement of couple and partner testing is a key component of the P4P program for HTC in Rwanda. While individual HTC is recognized as the necessary gateway for HIV/AIDS treatment, the prevention benefits of individual HTC remain under discussion (Denison et al., 2008). Joint couple or partner testing on the other hand appears to have stronger prevention benefits, especially in the case of discordant couples (Allen et al., 2003; Cohen et al., 2011). However, despite the apparent importance of couple testing for treatment and prevention purposes, there have been few successful experiences of HTC programs reaching couples (Padian et al., 1993; Painter, 2001). Furthermore, recent evidence on the effectiveness of ART for prevention of HIV transmission among couples makes this a key intervention of prevention programs in generalized epidemic countries (Dodd et al., 2010; El-Sadr et al., 2010; Wagner et al., 2010). Recent evidence on the prevention effectiveness of ART points to a 95% protection rate among discordant couples (Cohen et al., 2011). Our results show that P4P is an effective intervention to target discordant couples for HTC.

Our findings contribute to the growing evidence base that paying health facilities for performance is a feasible and effective method for improving health system performance. More importantly, the role of incentives in P4P is key. Because the comparison facilities' regular budgets were increased by an amount equal to the P4P payment to the treatment group, we were able to isolate the P4P incentive effect from the resource effect. This implies that the same results could not have been achieved by simply increasing the amount of resources without the incentives.

References

Allen S, Meinzen-Derr J, Kautzman M, Zulu I, Trask S, Fideli U, Musonda R, Kasolo F, Gao F, Haworth A. Sexual behavior of HIV discordant couples after HIV counseling and testing. AIDS 2003; 17(5); 733-740.

Alshamsan R, Majeed A, Car J, Millett C. Impact of pay for performance on inequalities in health care: systematic review. Journal of Health Services Research & Policy 2010; 15(3); 178-184.

Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. The Lancet 2011; 377(9775); 1421-1428.

Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics. Rapport annuel du TRAC 2006. Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics: Kigali; 2007.

Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics. 2008 Annual Report. Center for Treatment and Research on AIDS Malaria Tuberculosis and Other Epidemics: Kigali; 2008.

Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JHS, Godbole SV, Mehendale S, Chariyalertsak S, Santos BR, Mayer KH, Hoffmann IF, Eshleman SH, Piwowar-Manning E, Wang L, Makhema J, Mills LA, de Bruyn G, Sanne I, Eron J, Gallant J, Havlir D, Swindells S, Ribaudo H, Elharrar V, Burns D, Taha TE, Nielsen-Saines K, Celentano D, Essex M, Fleming TR, for the HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. The New England Journal of Medicine 2011; 365(6); 493-505.

Denison JA, O'Reilly KR, Schmid GP, Kennedy CE, Sweat MD. HIV voluntary counseling and testing and behavioral risk reduction in developing countries: A meta-analysis, 1990-2005. AIDS and Behavior 2008; 12(3); 363-373.

Dodd PJ, Garnett GP, Hallett TB. Examining the promise of HIV elimination by 'Test and Treat' in hyper-endemic settings. AIDS 2010; 24(5); 729-735.

Eichler R, Levine R. Performance incentives for global health: Potential and pitfallss. Center for Global Development: Washington, D.C.; 2009.

El-Sadr WM, Affrunti M, Gamble T, Zerbe A. Antiretroviral Therapy: A Promising HIV Prevention Strategy? Journal of Acquired Immune Deficiency Syndromes 2010; 55(Suppl 2); S116-121.

Gertler PJ, Vermeersch C. Using Performance Incentives to Improve Health Outcomes. World Bank Policy Research Working Paper No. 6100. The World Bank; 2012.

Glick P. Scaling up HIV voluntary counseling and testing in Africa: What can evaluation studies tell us about potential prevention impacts? Evaluation Review 2005; 29(4); 331-357.

Health Development and Performance. Rapport d'Enquête de Contre vérification par la Communauté dans les districts de Nyamasheke, Nyanza, Nyaruguru et Rulindo. 2008.

Institut National de la Statistique du Rwanda (INSR), ORC Macro. Rwanda Demographic and Health Survey. INSR and ORC Macro: Calverton, Maryland, USA; 2006.

Meessen B, Soucat A, Sekabaraga C. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? Bulletin of the World Health Organization 2011; 89(153-156.

Miller G, Luo R, Zhang L, Sylvia S, Shi Y, Foo P, Zhao Q, Martorell R, Medina A, Rozelle S. Effectiveness of Provider Incentives for Anaemia Reducation in Rural Child: A Cluster Randomized Trial. British Medical Journal 2012; 345(e4809.

MINALOC (Ministry of Local Government). Making decentralized service delivery work in Rwanda: Putting the people at the center of service provision, Policy Note. Government of Rwanda: Kigali; 2006.

Ministère de la Santé République du Rwanda. Guide de l'Approche Contractuelle. Modules de Formation des Acteurs Concernés pour le Démarrage PBF dans Rwanda. 2006.

Ministère de la Santé République du Rwanda. Module de référence de formation continue en planification familial. A l'usage des formateurs, superviseurs et prestataires au niveau des formations sanitaires. 2008.

Olken BA, Onishi J, Wong S. Should Aid Reward Performance? Evidence from a field experiment on health and education in Indonesia. NBER Working Paper No. 17892. 2012.

Padian NS, O'Brien TR, Chang YC, Glass S, Francis DP. Prevention of heterosexual transmission of Human Immunodeficiency Virus through couple counseling. Journal of Acquired Immune Deficiency Syndromes 1993; 6(1043-1048).

Painter TM. Voluntary counseling and testing for couples: a high-leverage intervention for HIV/AIDS prevention in sub-Saharan Africa. Social Science and Medicine 2001; 53(2001); 1397-1411.

Peabody J, R Shimkhada R, S Quimbo, J Florentino, MF Bacate, C MuCulloch, O Solon. 2011. Financial Incentives and Measurement Improved Physician's Quality of Care in the Philippines. Health Affairs. 30: p773

Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J, Young D. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database of Systematic Reviews 2011; 9(CD008451.

Thornton RL. The demand for, and impact of, learning HIV status. American Economic Review 2008; 98(5); 1829-1863.

UNAIDS. AIDS epidemic update: November 2011. UNAIDS: Geneva, Switzerland; 2012.

United Nations. The Millennium Development Goals Report 2011. United Nations: New York; 2011.

Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic Review: Effects, Design Choices, and Context of Pay-for-Performance in Health Care. BioMed Central Health Services Research 2010; 10(247-260).

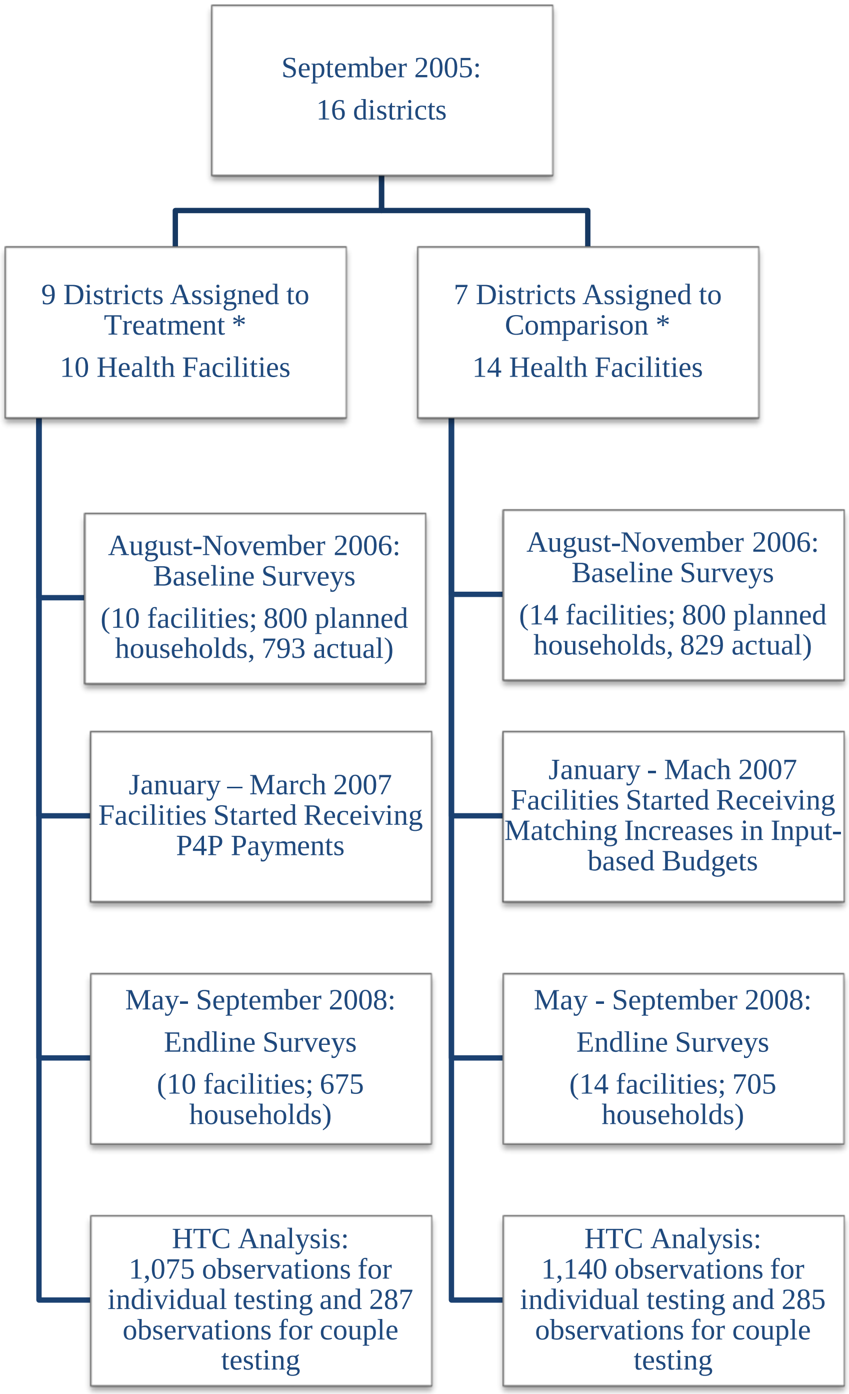
Vermeersch C, de Walque D, Sturdy J, Gertler P. Performance Based Contracting in Rwanda: Evaluation Design and Methods. The World Bank: Washington, D.C.; 2010.

Wagner BG, Kahn JS, Blower S. Should we try to eliminate HIV epidemics by using a "Test and Treat" strategy? AIDS 2010; 24(5); 775-776.

Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. Cochrane Database of Systematic Reviews 2011; 2(Art. No: CD007899).

World Health Organization. Towards universal access: scaling up priority HIV/AIDS interventions in the health sector. Progress Report 2009. World Health Organization: Geneva, Switzerland; 2009.

Figure 1: Experimental Design



* The originally planned evaluation consisted of 18 Phase I health facilities and 18 Phase II health facilities from 7 and 7 districts, respectively. Prior to implementation of the baseline survey, the administrative district boundaries were redrawn in the context of a decentralization effort. As a result, some of the experimental areas were combined with areas that already had NGO P4P schemes. Because P4P could not be "removed" from health facilities that were already implementing the system, and because P4P was managed at the district level, the GoR required that all facilities within those new districts be in the first phase

(treatment) of the rollout. This led the evaluation team to switch the assignment of treatment and comparison for eight districts from four blocks, as well as add one block to the sample.

Table 1: Output indicators and unit payments by P4P for HIV services

	Service	Quantity indicators for HIV	Amount paid by P4P per case (US$)
1	HTC	Number of clients tested for HIV at the HTC center	0.92
2	HTC / PMTCT	Number of couples/partners tested during the reporting month	4.59
3	PMTCT	Number of HIV+ pregnant women on ARV treatment during labor	4.59
4	PMTCT	Number of infants born to HIV+ mothers tested	9.17
5	Care	Number of HIV+ patients who received CD4 test	4.59
6	Care	Number of HIV+ patients treated with co-trimoxazole each month	0.46
7	ARV	Number of new HIV+ adults on ARV treatment	4.59
8	ARV	Number of new HIV+ infants on ARV treatment	6.88
9	HIV Prevention	Number of HIV+ women on contraception	2.75
10	HIV Prevention	Total number of HIV+ patients tested for tuberculosis	2.75

Notes:

- P4P: Pay-for-Performance
- HIV: Human Immunodeficiency Virus
- HTC: HIV Testing and Counseling
- PMTCT: Prevention of Mother-To-Child Transmission (of HIV)
- ARV: Antiretroviral drug

Table 2: Health facility baseline (2006) characteristics

	Treatment		**Control**		**P-value**
	Mean	**SE**	**Mean**	**SE**	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Located in rural area	0.900	0.107	0.714	0.175	0.384
Is a district hospital	0.600	0.142	0.500	0.129	0.612
Facility is public (vs. assisted by the government)	0.400	0.252	0.429	0.174	0.927
Catchment population	135928	20229	111014	20150.65	0.402
Number of staff per facility	60.800	16.435	55.000	7.620	0.755
Number of staff per 10 000 population	6.006	0.862	7.208	1.005	0.383
Log Total Expenditures (2006)	17.432	0.321	17.832	0.303	0.384
Log Total Expenditures (2008)	18.338	0.744	18.676	0.315	0.684
Medical personnel budget share	0.459	0.047	0.482	0.029	0.685
Medical supply budget share	0.282	0.043	0.264	0.039	0.759
Non-medical budget share	0.259	0.037	0.257	0.020	0.966
Total number of Health Facilities	**10**		**14**		

Note: Standard errors (SE) were cluster-adjusted using districts as clusters. P-Value is for the difference between the treatment and control groups.

Table 3: Baseline (2006) characteristics of adults (>15 years) not identified as HIV patients

	All					**Not living in couple**					**Living in couple**				
	Treatment (N=438)		**Control (N=445)**			**Treatment (N=217)**		**Control (N=215)**			**Treatment (N=221)**		**Control (N=230)**		
Variable	**mean**	**SE**	**mean**	**SE**	**p**	**mean**	**SE**	**mean**	**SE**	**p**	**mean**	**SE**	**mean**	**SE**	**p**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Female	0.628	0.039	0.582	0.034	0.391	0.682	0.046	0.661	0.042	0.736	0.575	0.048	0.509	0.043	0.330
Age	34.332	1.191	35.984	1.025	0.318	30.899	1.751	30.498	1.541	0.867	37.703	0.772	41.112	0.765	0.011
Education															
No education	0.292	0.027	0.283	0.025	0.809	0.272	0.029	0.200	0.029	0.110	0.312	0.050	0.361	0.044	0.481
primary	0.616	0.020	0.605	0.021	0.693	0.645	0.055	0.651	0.048	0.936	0.588	0.050	0.561	0.044	0.690
secondary or higher	0.091	0.028	0.112	0.024	0.579	0.083	0.044	0.149	0.038	0.284	0.100	0.030	0.078	0.026	0.606
Marital status															
married	0.505	0.039	0.517	0.034	0.819	-	-	-	-	-	-	-	-	-	-
divorced/widow	0.201	0.039	0.180	0.033	0.689	0.406	0.077	0.372	0.066	0.749	-	-	-	-	-
never married	0.294	0.043	0.303	0.036	0.878	0.595	0.077	0.628	0.066	0.749	-	-	-	-	-
Log household asset value	11.915	0.310	11.919	0.253	0.992	11.742	0.379	11.621	0.321	0.813	12.084	0.263	12.197	0.223	0.750
sexual activity															
never had sex	0.192	0.042	0.160	0.035	0.572	0.378	0.086	0.312	0.072	0.566	-	-	-	-	-
ever had sex but not in past 12 month	0.358	0.047	0.391	0.040	0.608	0.530	0.068	0.558	0.059	0.761	0.190	0.037	0.235	0.033	0.388
had sex past 12 months	0.450	0.032	0.449	0.029	0.994	0.092	0.023	0.130	0.022	0.258	0.801	0.035	0.748	0.032	0.288
partner of an HIV patient	0.155	0.029	0.173	0.025	0.652	-	-	-	-	-	0.308	0.073	0.335	0.060	0.779
Ever been tested	0.580	0.018	0.539	0.020	0.164	0.447	0.050	0.391	0.045	0.425	0.710	0.023	0.678	0.026	0.376

Note: Standard Errors (SE) were cluster-adjusted using districts as clusters. P-Value is for the difference between treatment and control groups.

Table 4: Baseline (2006) characteristics of adults (>15 years) not identified as HIV patients, living with their sexual partners and who self-reported having had sex in the 12 months preceding the survey

	All respondents				
	Treatment (N=179)		Control (N=180)		
Variable	**mean**	**SE**	**mean**	**SE**	**p**
	(1)	(2)	(3)	(4)	(5)
Female	0.559	0.060	0.467	0.052	0.272
Age	35.978	0.915	39.150	0.865	0.030
Education					
No education	0.263	0.070	0.311	0.059	0.605
primary	0.643	0.065	0.611	0.056	0.723
secondary or higher	0.095	0.037	0.078	0.032	0.731
Log household asset value	11.948	0.221	12.099	0.196	0.621
Partner of an HIV patient	0.374	0.079	0.356	0.066	0.860
Ever been tested for HIV	0.788	0.033	0.711	0.033	0.130
Partner has been tested for HIV	0.821	0.047	0.739	0.042	0.218
Couple has been tested for HIV	0.721	0.039	0.650	0.038	0.223

Note: Standard Errors (SE) were cluster-adjusted using districts as clusters. P-Value is for the difference between treatment and control groups.

Table 5: Estimated impact of PBF on HIV Testing and Counseling at the individual level

	All	**In couple**	**Not in couple**	**Not in couple and ever had sex**
	(1)	(2)	(3)	(4)
$\boldsymbol{\beta}^{*}$ **SE**	0.061 (0.036)	0.102 (0.033)	-0.004 (0.059)	-0.050 (0.069)
P-value	0.099	0.005	0.952	0.481
$\%\Delta$ ***	10.57%	14.52%	-0.90%	-9.21%
N	2,215	920	1,295	683

Note: β^{**} is the estimated effect of P4P controlling for year, and respondent's characteristics including age, gender, age, years of schooling, and log household wealth. Standard Errors (SE) were cluster-adjusted using districts as clusters and all models used a health facility fixed effect. P is the p-value for the difference between treatment and control groups; and $\%\Delta^{***} = \left(\beta/_{Control\ mean}\right) * 100$, where the baseline mean equals the mean of the dependent variable for the control group at endline (2008).

Table 6: Estimated impact of PBF on HIV Testing and Counseling at the couple level

	All	**Discordant couples**	**Non discordant couples**
	(1)	(2)	(3)
$\boldsymbol{\beta}^{**}$ **(SE)**	0.086 (0.050)	0.147 (0.064)	0.072 (0.064)
P-value	0.095	0.032	0.277
$\%\Delta^{***}$	12.01%	18.14%	10.98%
N	572	229	343

Note: β^{**} is the estimated effect of P4P controlling for year, and respondent's characteristics including age, gender, age, years of schooling, and log household wealth. Standard Errors (SE) were cluster-adjusted using districts as clusters and all models used a health facility fixed effect. P is the p-value for the difference between treatment and control groups; and $\%\Delta^{***} = \left(\beta/_{Control\ mean}\right) * 100$, where the baseline mean equals the mean of the dependent variable for the control group at endline (2008).